

Available online at www.sciencedirect.com**ScienceDirect**

Procedia Economics and Finance 10 (2014) 314 – 323

Procedia
Economics and Finance

www.elsevier.com/locate/procedia7th International Conference on Applied Statistics

GROWTH MODELLING OF METROPOLITAN PERFORMANCE INDICATORS.

An application by means of R software

Karima Kourtiti^{a*}, Miruna Mazurencu Marinescu^b, Peter Nijkamp^a^a *Department of Spatial Economics, VU University, Amsterdam*^b *Department of Statistics and Econometrics, The Bucharest University of Economic Studies*

Abstract

The current study aims at explaining the integrated performance of metropolitan areas in our world by using a new approach to the field, namely advanced growth models, which are an important variation of a much broader technique known as multi-level modeling. Extensive data bases on 35 world cities were derived from the so-called GPCI-data base created by the Institute of Urban Research of the Japanese Mori Memorial Foundation (2012), which served as input data for the proposed models. An illustration in R software was also carried out.

© 2014 Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Selection and peer-review under responsibility of the Department of Statistics and Econometrics, Bucharest University of Economic Studies.

Keywords: metropolitan performance; growth modelling, GPCI database, R software.

1. Introduction

The development of urban systems and metropolitan areas is critically dependent on various megatrends of a global nature. First, the change in the world population will likely amount to approx. 2-3 percent growth per annum in the few decades to come (see UN 2008). Consequently, our planet will most likely have to accommodate at least 9 billion people by the year 2050.

A second megatrend is the likely unequal spread of these rising numbers of people. It is forecasted that there will likely be an increasing geographic imbalance in the spatial dispersion of the world population, with a rapid rate of

* Corresponding author.
E-mail address: k.kourtiti@vu.nl.

increase in major continents like Latin America, Africa and Asia, but with a modest rate of increase, and even, a stable development in other parts of the world (in particular, Europe). Some countries like Japan or France may even show a population decline.

A third megatrend is likely to be a rise in urbanisation rates, especially in those countries that may face a fast population rise. World-wide, we observe that more and more people move to urban areas, to the extent that for the first time in human history, more than 50 percent of the world population is now living in urban areas (see Kourtit 2013). But these urbanisation rates show an increasing diversity, which is rapidly rising in megacities (with more than 10 mln inhabitants) and in metropolises (in general, urban agglomerations, with more than 1 mln inhabitants) (see Angotti 1996; McNeill 1999).

Metropolises have become the common settlement pattern all over the world, ranging from small agglomerations (e.g., Brussels) to megacities (e.g., Sao Paulo) (see Turok 2004). It should be added that in many cases such metropolises are not isolated islands, but are part of a broader network of interconnected metropolitan agglomerations. Joining forces among such urban regions reinforces the agglomeration benefits that are at the heart of any city formation (see Docherty et al. 2004; Ewen and Hebbert 2007; Kern and Bulkeley 2009; Neal 2012). Such intra-and inter-metropolitan externalities may be related to knowledge creation and transfer, innovative milieus, open and flexible labour markets, effective environmental and energy policy, joint marketing efforts etc.

Similar developments can be observed in poly-nuclear metropolitan developments, with a satellite system of suburban areas, (as opposed to a “Star” pattern that characterises older metropolis), edge cities and new towns around a central core area (Ile de Paris, GTA, LA etc.). All such geographically and functionally integrated settlement patterns serve to increase the efficiency and productivity – and hence the growth and competition potential – of urban agglomerations. But it is an as yet unanswered question which metropolises in our world are to be seen as the most successful ones – in other words, which ones would produce the highest performance – and why. This calls for a solid and empirical comparison of performance indicators – and their background factors – of major cities in our world.

This paper aims to advance knowledge on critical performance conditions of metropolises in our world by carrying out a statistical analysis of performance indicators for about 35 world cities. Extensive data-bases for these cities are derived from the so-called GPCI-data base constructed by the Institute of Urban Research of the Japanese Mori Memorial Foundation (2012).

2. Literature Review and Hypotheses

Scale advantages are the ‘raison d’être’ of growing business firms and forms the cornerstone for a contemporary global business. ‘Big size’ is not a human ideal, but an economic necessity to survive in a competitive environment. Increased productivity – supported by innovative behaviour – are (or represent) the critical success factor; to gain a strong competitive position in business life, and propels a permanent drive to do things better and to expand markets and market shares.

The above Neo-Schumpeterian interpretation of entrepreneurial dynamics holds essentially also for the life of cities. Cities are born and may grow, in population, in attractiveness or in wealth. Some cities conquer a stable position – sometimes even at a structurally low welfare level –, others may shrink – sometimes on a temporary basis as a result of unfavourable headwinds (Detroit, e.g.) –, and again others may gain an increasingly strong profile. The latter category is on a rising edge. Many cities in our world grow in population size, and several also in prosperity. But irrespective of urban economic welfare, most cities on our planet appear to grow in population numbers. They are transformed into large urban agglomerations (metropolises), while many of these big cities – especially in the developing world – grow into mega-cities. Our planet is gradually moving towards an urban planet (see Sassen 1991).

City growth all over the world is rather heterogeneous, however. To some extent, one may compare the evolution of cities with that of business firms. The laws of competition stipulate that firms try to achieve an increasingly larger share of their relevant markets. This strategy is also valid for cities in an open world. Such cities have to acquire and retain the patronage of (potential and actual) residents, visitors and business life. They have to attract various (selected) classes of such important economic agents to their territory. Consequently, a successful city is a city that is able to attract a significant share of the international market of urban agents and stakeholders. Thus, urban

evolution calls for an endogenous explanation of driving forces of city dynamics, both the impact of available urban attractiveness indicators and the implementation of appropriate urban governance measures. The scientific, mainly quantitative estimation and evaluation of the various factors that determine the socio-economic performance of large cities is called metropolitan performance analysis (MPA).

MPA is not a single and unambiguous research tool, but mainly a set of quantitative assessment instruments to map out the drivers of the performance of urban agglomerations. An overview of several of these tools can be found in a recent study by Kourtit (2013) which offers an empirical illustration of several assessment tools, such as: self-organizing mapping models, data envelopment analysis (DEA) models, rough set analysis, and multi-criteria analysis.

The information needed to perform such a comparative benchmark analysis for several cities may originate from several statistical sources, in particular:

- generally accessible statistical data on cities
- specific survey-based indicators on urban performance
- perceptions of stakeholders on relevant urban items
- quantitative assessments of urban attractiveness from expert opinions.

The recent literature on urban studies offers a wealth of contributions on comparable experiments regarding the economic, social, entrepreneurial, innovative or creative profile of cities. This has led to a variety of general ranking studies of global cities, world cities, metropolitan agglomerations etc, while also various sectoral or functional ranking studies have been performed, e.g. on tourist cities, business cities, financial cities, artistic cities and the like.

A noteworthy study was undertaken by Grosveld (2000) who applied the concept of Porter's (1990) competitive advantage theory to a comparison of leading cities in the world. His study is based on an extensive analysis of perceptions of 'city makers', from different socio-economic clusters: performing arts, hospitality, real estate and architecture, international trade and transport, corporate services, academia, museums, media, international organizations, multinationals, and finance. The information on these perceptions stems from an extensive and focussed survey questionnaire approach among some 85 leading world cities: 31 in Europe, 19 in North America, 16 in Asia, 9 in Latin America, 5 in Africa, and 3 in the Middle East. His findings offer a wealth of detailed insights into the functional-sectoral rank order of various cities in our world.

An important question – after the great many exploratory studies in the past decades – is of course whether an explanatory model can be designed that would offer a valid social linkage from a set of explanatory background factors to the performance indicator(s) of a given city or a set of cities. This would call for an econometric-statistical analysis of detailed and standardized data covering a large number of cities, while taking into account the specific geographic characteristics of such cities (e.g., developed versus developing world, city size class etc.). Fortunately, in the framework of our study, we had access to a rather comprehensive and very detailed, multi-level data-base relating to 35 metropolises of our world.

3. Methodological Aspects

Cities are engines of economic power but also nodes in global networks. They need each other, but are also each other's competitors. The combination of internal strength and external orientation determines the growth potential and economic position of cities (see Neal 2012).

Cities operate in an international playing field and, hence, their socio-economic performance may show much variation. The question is then: why do some cities outperform others? This idea formed the basis of the creation of the above-mentioned GPCI database. This database contains extensive information – in numerical form – of many world cities which are evaluated and ranked according to their 'magnetism', i.e. their competitive power to attract creative people and business enterprises from all over the world. This open access database is carefully validated through field visits and in-depth reports. It contains a wealth of multi-annual data on major critical indicators – and a very detailed list of sub-indicators – for economic strength of the relevant cities contained in the database. At present, this data system has accurate information on 35 world cities ranging from New York to Istanbul, and from

Tokyo to Geneva. This extensive GPCI database offers also the possibility for benchmarking of each individual city, in terms of strength and weakness regarding each individual performance indicator.

Thus, the GPCI aims to offer systematic and comparative information on the comprehensive economic position of major cities in the world, and it does so by focussing on a wide variety of functions performed by the cities under consideration. For each individual city, 6 main classes of functions were carefully mapped out and numerically assessed, viz. economy, research and development, cultural interaction, liveability, environment and accessibility. In addition, the importance of these indicators was carefully assessed by 5 distinct groups of stakeholders, viz. managers, researchers, artists, visitors and residents.

The purpose of this paper is how to address these above mentioned substantive research questions with multilevel modelling and then to briefly illustrate the proposed models in R. Multilevel analysis (or modelling) is a term used to describe a set of analyses also referred to as multilevel random coefficient models or mixed-effects models (Bryk and Raudenbush 1992; Kreft and De leeuw 1998; Snijders and Bosker 1999).

Usually, the definition of multilevel modelling reflects a wide range of interrelated multilevel topics (see also Klein and Kozlowski 2000), like within-group agreement and reliability, contextual OLS models, covariance theorem decomposition, Random Coefficient Modelling and Random Group Resampling. For the purpose of the current paper, we will restrict our analysis to the application of multilevel random coefficients models by using the GPCI data set gathered from the Mori Memorial Foundation Data Base between 2009 and 2011.

Firstly, PROC MIXED in SAS software was used in order to determine which city function is statistically significant for the various factors: managers, researchers, artists, visitors and residents.

The general model has the following expression (fixed effect model):

$$y_{it} = \alpha + X'_{it}\beta + v_{it} \quad (1)$$

where y_{it} is the dependent variable, X_{it} is a matrix of explanatory variables, $v_{it} \sim IID(0, \sigma_v^2)$; the index i refers to a city i , while the index t refers to the time-period.

Model (1) was estimated having as dependent variable the score for the various actors: managers, researchers, artists, visitors and residents, by using the following code lines:

```
proc mixed data=sc; class year;
model depvar=Economy RD CI Liv Env Acc /solution ;
random intercept/subject=year;
run;
```

Table 1. Managers- Fixed Effects

		Solution for Fixed Effects				
	Effect	Estimate	Standard Error	DF	t Value	Pr > t
Manager	Intercept	95624.000	59646.000	2	1.600	0.250
	Economy	0.656	0.052	96	12.650	<.0001
	RD	-0.108	0.031	96	-3.450	0.001
	Cultural Integration (CI)	0.049	0.041	96	1.210	0.231
	Livability (Liv)	-0.046	0.058	96	-0.790	0.431
	Environment (Env)	0.029	0.035	96	0.810	0.417
	Accessibility (Acc)	0.101	0.056	96	1.800	0.075

The only coefficient that is statistically significant is of the factor/dimension **Economy**, which was to be expected for the class of Managers (see Table 1).

Table 2. Researchers- Fixed Effects

Solution for Fixed Effects						
	Effect	Estimate	Standard Error	DF	t Value	Pr > t
Researcher	Intercept	-23146.000	59729.000	2	-0.390	0.736
	Economy	0.163	0.052	96	3.160	0.002
	RD	0.508	0.031	96	16.280	<.0001
	Cultural Integration	0.094	0.041	96	2.310	0.023
	Livability	0.091	0.057	96	1.590	0.116
	Environment	0.031	0.035	96	0.880	0.382
	Accessibility	0.024	0.056	96	0.430	0.670

For the researchers, the coefficients for Economy and R&D are statistically significant at a 5% significance level (see Table 2).

Table 3. Actors –Fixed Effects

Solution for Fixed Effects						
	Effect	Estimate	Standard Error	DF	t Value	Pr > t
Artist	Intercept	-74189.000	71052.000	2	-1.040	0.406
	Economy	-0.096	0.076	96	-1.260	0.210
	RD	0.136	0.046	96	2.980	0.004
	Cultural Integration	0.378	0.060	96	6.340	<.0001
	Liveability	0.513	0.084	96	6.090	<.0001
	Environment	-0.017	0.051	96	-0.340	0.735
	Accessibility	0.033	0.082	96	0.400	0.687

For the Artists, Cultural Integration and Liveability are statistically significant at a 5% significance level, with Livability being even more important than Cultural Integration (see Table 3).

Table 4. Visitors-Fixed Effects

Solution for Fixed Effects						
	Effect	Estimate	Standard Error	DF	t Value	Pr > t
Visitor	Intercept	6824.000	60126.000	2	1.130	0.374
	Economy	0.235	0.056	96	4.200	<.0001
	RD	-0.064	0.034	96	-1.900	0.061
	Cultural Integration	0.406	0.044	96	9.240	<.0001
	Liveability	0.158	0.062	96	2.530	0.013
	Environment	-0.048	0.038	96	-1.270	0.206
	Accessibility	0.082	0.061	96	1.360	0.176

Visitors value more Cultural Integration followed by Economy and Liveability, these three factors being the only significant ones at a 5% significance level (see Table 4).

Table 5. Residents- Fixed Effects

		Solution for Fixed Effects				
	Effect	Estimate	Standard Error	DF	t Value	Pr > t
Resident	Intercept	-85251.000	69776.000	2	-1.220	0.346
	Economy	0.251	0.063	96	3.980	0.000
	RD	0.179	0.038	96	4.710	<.0001
	Cultural Integration	-0.065	0.049	96	-1.320	0.190
	Liveability	0.520	0.070	96	7.420	<.0001
	Environment	0.109	0.043	96	2.560	0.012
	Accessibility	0.140	0.068	96	2.060	0.042

As for the residents, Liveability matters most, but Economy, R&D, Accessibility and Environment are the other statistically significant dimensions at a 5% significance level (see Table 5).

4. Model Specification

With the data in univariate form, it is possible to examine visually whether or not there are identifiable patterns between time and the outcome (city total score). The commands below use the lattice package to produce the plot of all 35 cities (see Figure 1).

```
> library(lattice)
> xyplot(MULTDV~TIME|as.factor(city), data=UNIV.datacities[1:105,], type=c("p", "r", "g"), col="blue",
col.line="black", xlab="TIME", ylab="Score")
```

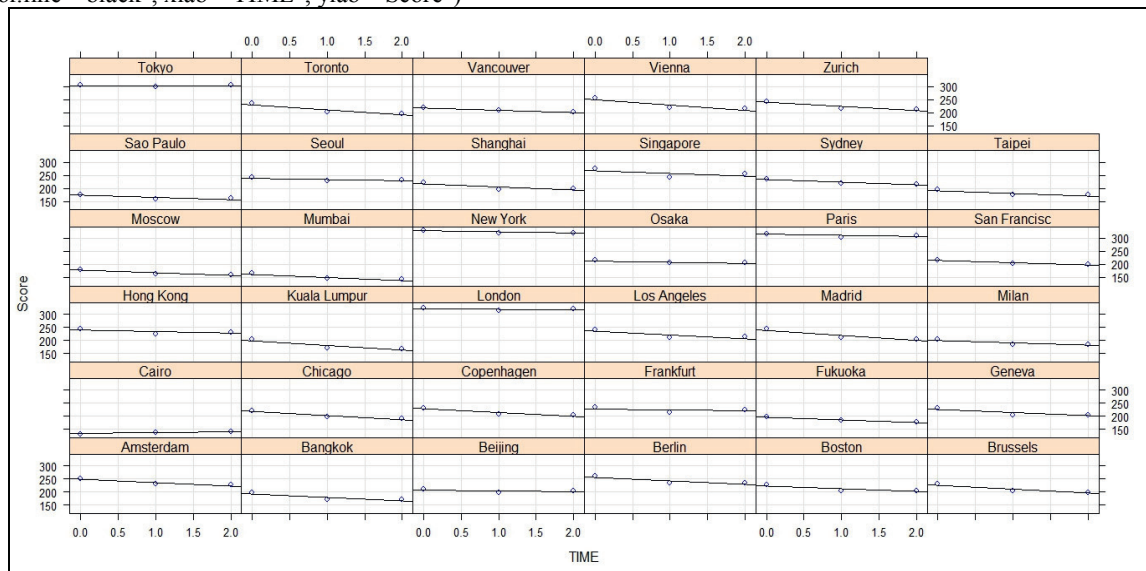


Fig. 1.City variability across time

From this plot, we observe that there is variability both in overall levels of the total score and in the way it changes over time. The goal in our growth modelling exercise is to determine whether or not there are consistent

patterns in the relationship between time and the dependent variable (DV), which is the city total score. This will be highlighted in the subsequent subsections.

4.1. Examination of the DV

The first step in growth modelling is to examine the properties of the dependent variable. As in classical multilevel modelling, one begins by estimating a null model and calculating the ICC[†].

```
> null.model<-lme(MULTDV~1, random=~1|city, data=UNIV.datacities, na.action=na.omit,
control=list(opt="optim"))
> VarCorr(null.model)
city = pdLogChol(1)
```

Table 6. The null model

	Variance	StdDev
(Intercept)	1812.670	42.575
Residual	174.583	13.213

```
> 1812.6705/(1812.6705+174.5839)
[1] 0.9121482
```

The ICC associated with the total score is 0.9121. This indicates that 91.21% of the variance in any city score can be explained by the intrinsic properties of the city itself.

4.2. Model Time

The second step involves modelling the fixed relationship between time and the dependent variable. We shall begin by modelling a linear relationship and add progressively more complicated relationships such as quadratic, cubic, etc. In order to test if there is a linear relationship between time and the city total score, we regress the total score on time in a model with a random intercept.

```
> model.2<-lme(MULTDV~TIME, random=~1|city, data=UNIV.datacities, na.action=na.omit,
control=list(opt="optim"))
> summary(model.2)$tTable
```

Table 7. Model 2

Value	Std.Error	DF	t-value	p-value	
(Intercept)	228.292	7.384	69.000	30.918	0.000
TIME	10.044	1.034	69.000	-9.718	0.000

An examination of the fixed effects indicates that there is a significant linear relationship between time and the city total score, such that the total score increases by 10,04 each time period. Since the first time period was coded as 0, the intercept value in this model represents the average level of the total score in the first time period. Specifically, at the first time period the average city total score was found to be equal to 228.29.

More complicated time functions can be included in one of two ways – either through raising the time variable to various power levels, or by converting time into power polynomials. Below, both techniques are illustrated.

[†] The intraclass correlation (or the *intraclass correlation coefficient*, abbreviated ICC is a descriptive statistic that can be used when quantitative measurements are made on units that are organized into groups. It describes how strongly units in the same group resemble each other.

```
> model.2b<-lme(MULTDV~TIME+I(TIME^2), random=~1|city, data=UNIV.datacities, na.action=na.omit,
control=list(opt="optim"))
> summary(model.2b)$tTable
```

Table 8. Model 2b

Value	Std.Error	DF	t-value	p-value
(Intercept)	231.696	7.363	68.000	31.467
TIME	-30.469	2.728	68.000	-11.168
I(TIME^2)	10.213	1.311	68.000	7.792

```
>model.2c<-lme(MULTDV~poly(TIME,2), random=~1|city, data=UNIV.datacities, na.action=na.omit,
control=list(opt="optim"))
> summary(model.2c)$tTable
```

Table 9. Model 2c

	Value	Std.Error	DF	t-value	p-value
(Intercept)	218.248	7.311	68.000	29.851	0.000
poly(TIME, 2)1	-84.033	6.331	68.000	-13.273	0.000
poly(TIME, 2)2	49.332	6.331	68.000	7.792	0.000

Both models clearly show that there is a significant quadratic trend. We can conclude that time has a quadratic relationship with the total city score.

4.3. Model Slope Variability

A potential limitation with model 2 is that it assumes that the relationship between time and city total score is constant for all cities. Specifically, it assumes that each city score increases by 10,04 at each time period.

An alternative model that needs to be tested is one that allows slopes to randomly vary. Given the degree of variability in the graph above, a random slope model seems quite plausible with the current data. The random slope model is tested by adding the quadratic effect for time as a random effect.

Consequently we will update model.2 by adding a different random effect component and contrast model 2 and model 3.

```
> model.3c<-update(model.2, random=~TIME+I(TIME^2)|city)
> anova(model.2, model.3c)
```

Table 10. Model 3

Model	df	AIC	BIC	logLik	Test
model.2	1	4	900.415	910.954	-446.207
model.3c	2	9	892.240	915.952	-437.120

	L.Ratio	p-value
1 vs. 2	18.175	0.003

The results clearly indicate that a model that allows the slope between time and city total score to randomly vary fits the data better than a model that fixes the slope to a constant value for all cities.

4.4. Modeling Error Structures

The fourth step in developing the level-1 model involves assessing the error structures of the model. It is important to carefully scrutinize the level-1 error structures because significant tests may be dramatically affected if error structures are not properly specified. The purpose of this step is to determine whether the model fit improves by incorporating (a) an autoregressive structure with serial correlations and (b) heterogeneity in error structures.

Tests for an autoregressive structure (autocorrelation) are conducted by including the correlation option in lme. We shall therefore update model .3c and include a lag 1 autocorrelation as follows:

```
> model.4a<-update(model.3c, correlation =corAR1())
> anova(model.3c, model.4a)
```

Table 11. Model 4a

Model	df	AIC	BIC	logLik	Test
model.3c	1	9	892.240	915.952	-437.120
model.4a	2	10	894.240	920.587	-437.120

	L.Ratio	p-value
1 vs. 2	0.000	0.998

A model that allows for autocorrelation does not fit the data better than a model that assumes no autocorrelation.

We can further examine the degree to which the variance of the responses changes over time. A test of variance homogeneity is conducted by examining the variance of the city total score at each time point using the `tapply` command.

```
> tapply(UNIV.datacities$MULTDV, UNIV.datacities$TIME, var, na.rm=T)
```

Table 12. Test for variance homogeneity

0	2	1
1799.609	1857.982	2035.171

The analysis suggests that the variance of the city total score is increasing over time.

5. Conclusions

Firstly, we have tested which city function is statistically significant for the various actors: managers, researchers, artists, visitors and residents by using a fixed effect model in SAS software. For the managers, the dimension *Economy* is statistically significant; for the researchers, the coefficients for *Economy* and *R&D* are statistically significant at a 5% significance level; for the artists, *Cultural Integration* and *Livability* matter most; visitors value more *Cultural Integration* followed by *Economy* and *Livability*.

Secondly, in order to expand the analysis, growth models from the multi-level modeling methodology have been applied. The results indicate that 91.21% of the variance in any city score can be explained by the properties of the city itself.

An examination of the fixed effects indicates that there is a significant linear relationship between time and the city total score such that total score increases by 10.04 each time period. Since the first time period was coded as 0, the intercept value in this model represents the average level of the total score in the first time period. Specifically, at the first time period the average city total score was found to be equal to 228.29.

We can also conclude that time has a quadratic relationship with total score.

In addition, the results clearly indicate that a model that allows the slope between time and the city total score to randomly vary fits the data better than a model that fixes the slope to a constant value for all cities.

Finally, a model that allows for autocorrelation does not fit the data better than a model that assumes no autocorrelation.

These preliminary findings will lead to better articulate metropolitan development policies in order to increase the cities' total score. Thus, the cities will become more appealing to various stakeholders, while maintaining at the same time their competitive advantage.

Unfortunately, we have limited observation waves, but the current research will be further extended once we gain access to more data.

References

- Angotti, T., 1995. The Metropolis Revisited. *Futures*, 27 (6), 627-639.
- Docherty, I., Gulliver, S., & Drake, P., 2004. Exploring the Potential Benefits of City Collaboration. *Regional Studies*, 38 (4), 445-456.
- Ewen, S., & Herbert, M. (2007). European Cities in a Networked World during the Long 20th Century. *Environment and Planning C: Government and Policy*, 25, 327-340.
- Grosveld, H., 2000. *The Leading Cities of the World and their Competitive Advantages*. Thela Thesis, Amsterdam
- Institute of Urban Study, 2012. *Global Power City Index*. Mori Memorial Foundation, Tokyo.
- Kern, K., & Bulkeley, H., 2009. Cities, Europeanization and Multi-level Governance: Governing Climate Change through Transnational Municipal Networks. *Journal of Common Market Studies*, 47 (2), 309-322.
- Kourtit, K., 2013. *The New Urban World*. Ph.D Thesis, VU University, Amsterdam
- Mazurencu Marinescu M., Pele, D.T., 2012. Applying multilevel modelling for assessing business performance in software: case study--the companies operating in Romania's ICT sector, *Journal of International Business and Economics*, ISSN: 1544-8037, Volume 12, Number 3, 2012, pg. 99-102
- McNeill, D. 1999. Globalization and the European city. *Cities*, 16 (3), 143-147.
- Neal, Z., 2012. *The Connected City*. Routledge.
- Porter, M., 1990. *The Competitive Advantage of Nations*. MacMillan, London.
- Sassen, S., 1991. *The Global City*, Princeton University Press, Princeton
- Singer, J. D., 1998. Using SAS Proc Mixed to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, 24, pp. 323-355.
- Singer, J. D., and Willett, J. B., 2003. *Applied longitudinal data analysis*. New York: Oxford University Press.
- Snijders, A.B. Tom and Boskere, J. Roel, 2003. *Multilevel analysis, An introduction to basic and advanced multilevel modeling*. The Cromwell Press Ltd.
- Snijders, T. A. B. and Bosker, R. J., 1999. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage Publications.
- Statistical Computing Seminars, *Repeated Measures Analysis with R*. Institute for Digital Research and Education, UCLA. [online] Available at <http://www.ats.ucla.edu/stat/R/seminars/Repeated_Measures/repeated_measures.htm>
- Turok, I., 2004. Cities, Regions and Competitiveness. *Regional Studies*, 38 (9), 1069-083.